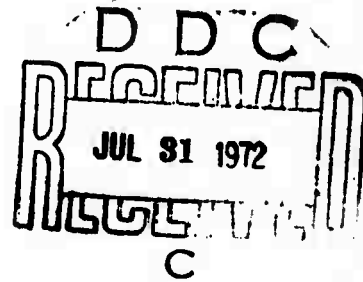# THE INTERACTION OF REASONING AND MEMORY ABILITIES WITH RULE-EXAMPLE
VS. DISCOVERY INSTRUCTION IN LEARNING AN IMAGINARY SCIENCE

C. Victor Bunderson
Paul F. Merrill
William P. Olivier

TECHNICAL REPORT NO. 3

August 1971

Co-Principal Investigators:
C. V. Bunderson and J. L. Dunham

THE UNIVERSITY OF TEXAS AT AUSTIN
Computer-Assisted Instruction Laboratory
Austin, Texas 78712

AD 745718

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)	2a. REPORT SECURITY CLASSIFICATION
Advanced Research Projects Agency Washington, D.C. 20301	*Unclassified*
	2b. GROUP

3. REPORT TITLE

THE INTERACTION OF REASONING AND MEMORY ABILITIES WITH RULE-EXAMPLE VS. DISCOVERY INSTRUCTION IN LEARNING AN IMAGINARY SCIENCE

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Technical Report No. 3, August 1971

5. AUTHOR(S) (First name, middle initial, last name)

C. Victor Bunderson, Paul F. Merrill, William F. Olivier

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
August 1971	22	13

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
N00014-67-A-0126-0006	
b. PROJECT NO. NR 154-282/8-13-68 Code 458	
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d.	Technical Report No. 3

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	U. S. Department of the Navy Office of Naval Research Washington, D. C.

13. ABSTRACT

I

DD FORM 1 NOV 65 **1473** (PAGE 1)

S/N 0101-807-6811

THE INTERACTION OF REASONING AND MEMORY ABILITIES WITH RULE-EXAMPLE

VS DISCOVERY INSTRUCTION IN LEARNING AN IMAGINARY SCIENCE

INTRODUCTION

One of the most significant discussions of the learning by discovery issue is found in the publication (Shulman & Keislar, 1966) of the proceedings of a conference on Learning by Discovery, sponsored by the Committee on Learning and the Educational Process of the Social Science Research Council. Although each paper presented at the conference was relevant to the learning by discovery issue, two of the papers were especially relevant to research on learning by discovery. Wittrock (1966) discussed the issues and problems associated with the learning by discovery hypothesis as proposed by Bruner (1961) and reviewed several representative studies in detail. He stated that it was very difficult to derive any general conclusions from the data because of the wide variety of dependent variables used and the lack of operationally defined treatments. He stressed that it is futile to expect one method of learning to be consistently superior to all others using a variety of subject matter and $\underline{S}$s from different populations. He suggested that future studies should take into account the history and individual differences of the $\underline{S}$s and that investigators should search for alternatives to the all-or-none position. He further recommended that treatments should be designed to vary only one element at a time and that the treatments should differ in a systematic and meaningful way.

Additional recommendations relevant to future studies on learning by discovery were made in Cronbach's (1966) conference paper. He argued that learning tasks used in future research should have psychological properties similar to those of educational subject matter. Thus, he discouraged the use of tasks in which the S-R linkage was arbitrary and recommended that the principles or rules of a learning task should fit into a system of supporting propositions. He also suggested that interactions of pupil characteristics with treatments on multiple outcomes should be investigated. He observed that most of the research on learning by discovery is confounded because of an experimental design dilemma. If each $\underline{S}$ in both the discovery and the expository treatment groups is trained to the same performance criterion, the discovery $\underline{S}$s may require more time and examples to reach the criterion. The treatments would therefore be confounded because of the differential content and time received by each $\underline{S}$. On the other hand, if the experimental design equates time and content for each $\underline{S}$, scores on retention and transfer measures may be confounded by nondiscoverers or overlearners.

1

It is impossible to follow all of the recommendations and to solve all the problems mentioned above in any single experimental study. However, the investigator should be aware of which supposed research requirements are satisfied and which have not been satisfied; then, the results of any study should be interpreted and qualified accordingly.

The purpose of the present study was to investigate the relationships between individual difference variables and discovery vs. non-discovery instructional treatments. Both Wittrock (1966) and Cronbach (1966) suggested that future studies should investigate the relationships between pupil characteristics and instructional treatments. Cronbach and Snow (1969) reported the results of a pilot study conducted by Jane Stallings which investigated the relationships between psycholinguistic and memory abilities and a phonics vs. look-say reading instruction treatments. The results showed that high ability $\underline{Ss}$ obtained high reading achievement scores when assigned to a look-say treatment, while low ability $\underline{Ss}$ obtained high reading achievement scores when assigned to a phonics treatment. Cronbach and Snow (1969) argued that "if these results can be substantiated, then the age-old battle about which of these reading methods is the 'one best way' seems a very hollow fight." Therefore, as suggested by Wittroch (1966), this study was conducted to provide an alternative to the all-or-none position on learning by discovery and search for relationships between treatments and individual differences.

Earlier studies by Bunderson (1967) and Dunham, Guilford and Hoepfner (1968) have demonstrated strong relationships between task performance and cognitive ability measures, while studies by Dunham and Bunderson (1969) and those reported by Cronbach and Snow (1969) have revealed that relationships between task performance and abilities may be altered by manipulating a task variable.

In a pilot study (Bunderson, Olivier, & Merrill, 1971), the feasibility of using an imaginary science adapted for presentation on the IBM 1500 computer-assisted instruction (CAI) system as a learning task for investigating learning by discovery variables was evaluated. By following a task analysis procedure outlined by Gagné (1961), a hierarchy of 13 rules for the science were developed, and a series of tabular displays to be generated by the computer was selected to serve as examples of each rule. The $\underline{Ss}$ were 51 students enrolled in science education classes at The University of Texas at Austin, who were randomly assigned to an example-only, or discovery, treatment group and a rule-example, or non-discovery, treatment group. All $\underline{Ss}$ received additional examples of a rule until they were able to pass constructed response test items on the rule. Before learning the task, all $\underline{Ss}$ were given a battery of cognitive tests designed to measure the abilities of induction, associative memory, and general reasoning. The groups differed only on the number of examples required

2

to learn the science, $F(1,49) = 9.087$, $p < .01$, with more examples being required by the example-only group. There were no significant differences between the groups on the post, retention, or transfer tests. The battery of cognitive tests were factor analyzed, and a two-factor Varimax solution was obtained yielding the factors of reasoning and associative memory. Regression analyses of the factor scores and the criterion measures were conducted. A significant disordinal interaction, $F(1,47) = 7.272$, $p < .01$, between the associative memory factor scores and the treatments, using number of examples as criterion, was found. This analysis revealed that Ss with high memory scores in the rule-example group required few examples, while Ss with low memory in the example-only groups required few examples. There were no significant interactions with the reasoning factor scores. In addition to the results cited above, this pilot study revealed that Ss required an excessive amount of time to learn the science, while their level of performance on the posttest was mediocre. Several of the rules were too difficult, and many Ss needed help and encouragement to complete the task. It was also found that the generation of the tabular displays, which served as examples of the rules, required an excessive amount of computer time and was distracting to the Ss. (The algorithm used to generate these displays used the integer arithmetic capabilities of Coursewriter II, and was very inefficient.) The generation of the tables also made it necessary to require Ss to make selected observations and record them in a workbook to focus their attention on relevant parts of a given table. Copying this information, of course, became quite tedious.

Based on the findings of the pilot study, the imaginary science and experimental procedure were revised, and the studies reported in this paper were subsequently conducted using the revised version of the science.

METHOD

Experiment I

Subjects

The 30 Ss who participated in this study were enrolled in two sections of a science education course at The University of Texas at Austin. The Ss were required to participate as a class assignment.

Ability Measures

Seven cognitive ability tests were selected from the Kit of Reference Tests for Cognitive Factors (French, Ekstrom, & Price, 1963) to measure the abilities expected to be related to task performance. The

Object-Number Test and the First and Last Names Test were selected to measure associative memory. Inductive reasoning was measured by the Letter Sets Test and the Locations Test. The Ship Destination Test, Necessary Arithmetic Operations Test, and Mathematics Aptitude Test were chosen to measure general reasoning.

Experimental Task, Materials, and Equipment

The learning task used in this study was a revised version of the imaginary science called the Science of Xenograde Systems. The use of an imaginary science assured that all Ss would have had no previous experience with the task, thus eliminating the necessity of pretesting and discarding Ss who might be familiar with the task content. With prior knowledge of the task principles held constant, the source of individual differences could, with greater confidence, be attributed solely to Ss' different abilities and learning styles. Since the principles of the science were interrelated and similar in structure and content to formal science topics, the recommendation by Cronbach (1966) that the experimental task should have psychological properties similar to those of subject matter taught in the schools was followed.

The science was conceived by Carl Bereiter at the Training Research Laboratory, University of Illinois, and was expanded and developed by David Merrill (1964). Merrill's version of the science was simplified and revised, and an instructional program for presenting the task on the IBM 1500 computer-assisted instruction system was designed according to the instructional design model developed by Bunderson (1969).

In the revised version of the science used in this study, a Xenograde System consists of a satellite which revolves around a nucleus. The satellite and nucleus contain particles called alphons. The laws and relationships between the components of the system as a function of time and initial conditions comprise the subject matter of the task. The terminal objective of the task requires the Ss to predict the state of a Xenograde System for each unit of time from time zero to a specified time when given the initial conditions of the system. Operationally, this consisted of filling in a table of readings called a Xenograde table. Ten rules for the revised science and the hierarchical sequence for presenting the rules were determined by an analysis of an efficient information-processing algorithm for performing the terminal behavior. A complete description of this analysis procedure and the design of the initial program may be found in Bunderson, Olivier, and Merrill (1971).

The materials used in the instructional program consisted of statements of ten rules, five examples for each rule, and five short constructed response tests for each rule. The examples were in the form

4

of partial Xenograde tables which demonstrated the activity of the Xeno-grade System components at several points in time. The short tests each contained three constructed response items which required Ss to apply the corresponding rule in order to make a correct prediction concerning the state of a Xenograde System at a given point in time.

The examples and test items were displayed on the cathode ray tube (CRT) of the IBM 1510 instructional terminal, while the Ss responded to the test items by means of a typewriter keyboard on the terminal. It should be noted that the examples for this revised version of the science were stored and displayed rather than generated by an algorithm as was done in the pilot study described earlier. Thus, the portions of the table irrelevant to the rule could be omitted from the display. The state-ments of the rules were displayed on the IBM 1512 image projector terminal. The instructional program was written in the Coursewriter II language and presented by the IBM 1500/1800 CAI system. The use of the CAI system made it possible to present and withdraw any display at random under program control, to record accurate display and response latencies, and to record and score each student response.

A printed instruction booklet was also provided which contained an introduction to the science, the purpose and justification of the course, instructions on reading Xenograde tables, and a treatment-specific explana-tion of the procedure for learning the task. Samples of the booklet, examples, rules, and test items may be found in Bunderson et al. (1971).

Post and Retention Tests

The post and retention tests were parallel forms which required Ss to fill in entries of a Xenograde table given the initial conditions of the system. These measures were presented on the CRT terminals. In order to prevent cumulative errors, Ss were given corrective feedback after making entries for a complete line of the table.

Transfer Task

The transfer task consisted of a booklet containing two Xenograde tables which served as examples of three higher-order rules of the science, not yet encountered by the Ss. The Ss were required to infer the rules from these examples and to apply the inferred rules in making predictions concerning the state of a Xenograde system at a given point in time. There were a total of nine test items with three items for each rule.

Procedure

In order to increase the interest and motivation of the Ss in the task, a short lecture was presented by E which explained the value of parti-cipation in the study and gave an introduction to CAI and ability by treatment

interaction studies. Following the lecture, all Ss were given the battery
of cognitive tests and were then instructed to schedule a two-hour session
at the CAI Laboratory to learn the Xenograde Science.

The Ss were randomly assigned to two treatment groups: an example-
only group and a rule-example group. All Ss were given the instruction
booklet before learning the science. The Ss in the rule-example group were
presented with a display of a statement of the first rule on the image pro-
jector. A Xenograde table which served as an example of the rule was sim-
ultaneously displayed on the CRT. After studying the rule and example,
the Ss responded to a three-item constructed response test requiring them
to predict certain values using the rule. If two of the three items were
answered correctly, S was given the next rule of the science; otherwise,
he was given another example of the same rule and another three-item test.
This sequence of new examples followed by response tests continued until
S had answered correctly two of the three items or had received five examples.
The task was completed after all ten rules of the science had been learned.
The Ss in the example-only group learned the science by the same basic pro-
cedure, except no statement of the rule was provided. The Ss in this group
were required to infer their own rule from the example(s) presented. Follow-
ing completion of the task, all Ss were given the posttest and were scheduled
to take the retention and transfer tests two weeks later. The results of
this study are reported below in connection with the results of a replication
experiment.

Experiment II

Since only 20 Ss completed all phases of the first experiment, a
second experiment was conducted to replicate the first one. The data from
the first experiment revealed that the revised version of the task was an
appropriate vehicle for studying ability by treatment interactions in
learning by discovery (Bunderson et al., 1971). Therefore, the same task,
materials, and procedure were used in Experiment II. An additional 53 Ss
from science education classes participated in this study. However, not
all Ss completed all phases of the study, and data from a few Ss were
lost because of computer malfunctions.

Results

Since Experiment II was an exact replication of Experiment I, the
results of both studies are presented in this section, along with the
results obtained when data from both studies were combined. In addition
to the scores obtained on the cognitive ability tests, the posttest,
retention test, and transfer test, data were obtained on the total number
of examples each S received.

6

The means and standard deviations for each criterion variable for Experiment I, Experiment II, and the combined data of both experiments are found in Table 1. The differences between the posttest means were not significant in either Experiment I or II. However, when the data were combined, a significant difference, $F(1,77) = 3.867$, $p < .05$, in favor of the rule-example group was obtained between the posttest means. Although the group retention test means were not significantly different in Experiment II, but they were in Experiment I, $F(1,16) = 4.588$, $p < .05$, and also when the data were combined, $F(1,65) = 8.331$, $p < .01$. In each case, the rule group means were greater than the example-only group means. There were no significant differences between the transfer test group means for either experiment or for the combined data. The groups differed on the number of examples required to learn the science [Experiment I: $F(1,28) = 11.983$, $p < .01$; Experiment II: $F(1,48) = 20.436$, $p < .001$; combined data: $F(1,78) = 30.926$, $p < .001$], with more examples required by the example-only group in both experiments and with combined data.

Separate factor analyses were conducted on the ability test scores from each experiment, and a third factor analysis was conducted on the ability tests scores from both experiments combined. In each case, a two-factor Varimax solution was obtained yielding the factors of reasoning and associative memory. Since the three solutions were very similar, only the factor matrix for the combined data is found in Table 2. The First and Last Names Test and the Object Number Test loaded on the associative memory factor, while the induction and general reasoning tests all loaded on the reasoning factor.

Regression analyses of the factor scores and the criterion measures were conducted for each experiment and for the combined data. A significant interaction, [Experiment I: $F(1,26) = 4.839$, $p < .05$; Experiment II: $F(1,46) = 11.158$, $p < .01$; combined data: $F(1,76) = 19,274$, $p < .001$] of reasoning factor scores with example-only vs. rule-example treatments using number of examples as criterion was found for both experiments and the combined data. Figure 1 is a plot of the regression lines for the combined data. (Corresponding plots for Experiment I and II were very similar.) Figure 1 shows that the number of examples required to learn the science has a high negative relationship to reasoning for $\underline{S}$s in the example-only group, while number of examples has little or no relationship to reasoning for $\underline{S}$s in the rule-example group.

There were no significant interactions between the treatments and memory factor scores using data from the separate experiments, but a memory-by-treatment interaction approached significance, $F(1,76) = 3.850$, $p < .10$, when the data were combined. Figure 2 is a plot of the regression lines for the combined data, and from this figure it can be seen that the relationship between memory and the treatments was very similar to that of reasoning (Figure 1).

Table 1

Group Means and Standard Deviations for Number of Examples, Posttest, Retention Test, and Transfer Test

Group	Number of Examples			Post Test			Retention Test			Transfer Test		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Experiment I												
Example-Only	14	14.0	3.6	14	88.8	6.3	10	91.9	5.3	10	4.6	1.8
Rule-Example	16	10.7	.84	16	92.4	7.0	8	96.5	2.3	10	5.6	1.7
Experiment II												
Example-Only	24	16.0	5.6	22	88.8	10.0	22	92.8	4.7	21	5.5	1.6
Rule-Example	26	10.8	1.2	27	92.8	9.2	27	95.4	4.1	25	5.2	2.2
Combined data from Experiments I & II												
Example-Only	38	15.3	5.1	36	88.8	8.8	32	92.5	4.9	31	5.2	1.8
Rule-Example	42	10.8	1.1	43	92.6	8.4	35	95.6	3.8	35	5.3	2.1

Table 2

Varimax Rotation Factor Matrix* on Combined Data

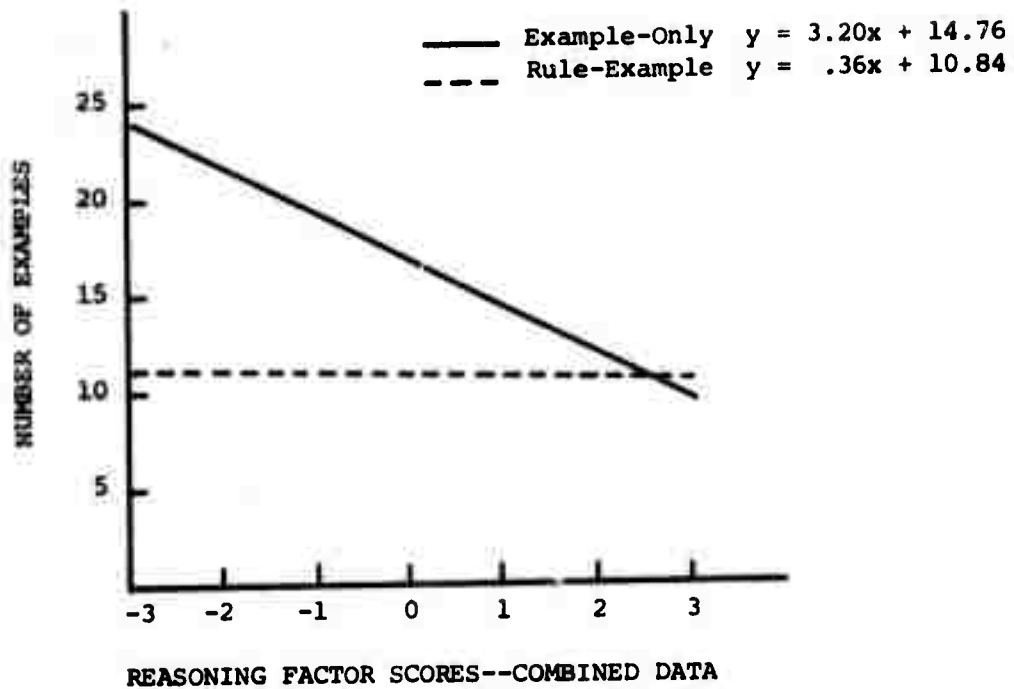| Tests	Factor Loadings	
	Reasoning Factor	*Associative Memory Factor*
Object Number (MA)	03	87
First Last Names (MA)	11	87
Locations (I)	66	11
Letter Sets (I)	70	16
Ship Destination (R)	87	09
Necessary Arithmetic Operations (R)	76	06
Mathematics Aptitude (R)	74	-11

*Decimal points omitted.

Figure 1.--Interaction of Reasoning Factor Scores and Treatments with Number of Examples as Criterion.
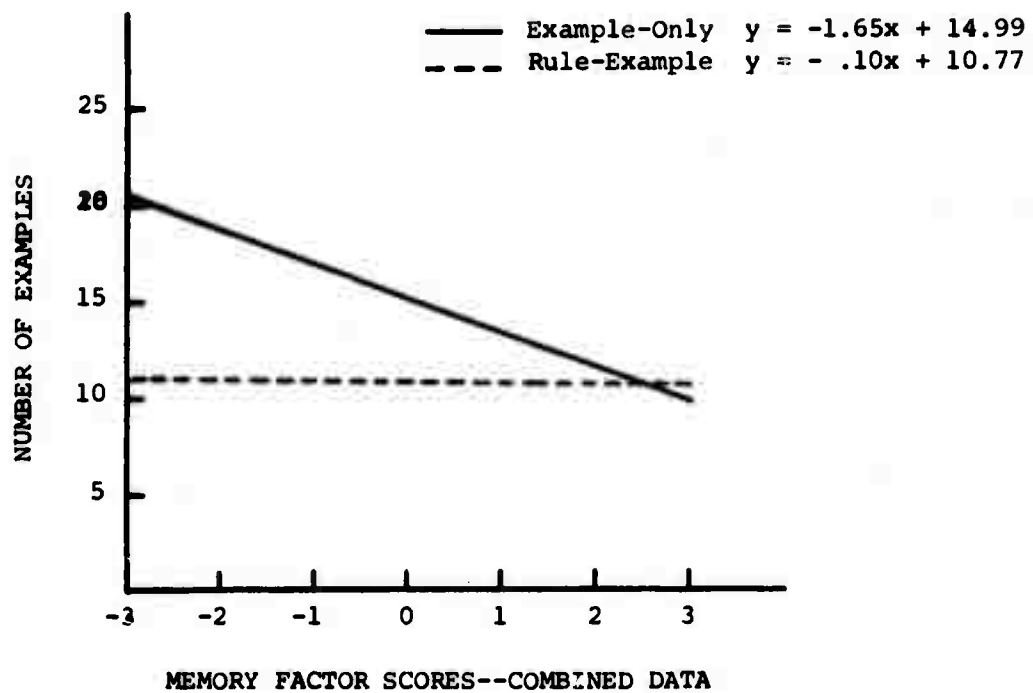
10

Figure 2.--Interaction of Memory Factor Scores and
Treatments with Number of Examples as Criterion.

11

DISCUSSION

The experimental procedure used in the two studies reported here required that all Ss reach a minimum criterion performance on each rule before they proceeded to learn the succeeding rule. By using this procedure, it was expected that both treatment groups would perform equally well on the terminal criterion measure. This expectation was supported when the data from the studies were analyzed separately, but an analysis of the combined data revealed a significant difference between the post-test means. The differences between retention test means were also not consistent across all three analyses. The retention test differences were consistently in the same direction, but only two of the three analyses revealed significant differences. The reason for these inconsistent findings on the post- and retention tests is not clear. However, the high scores obtained on these tests apparently were due to the corrective feedback which was given during the tests to prevent cumulative errors. This corrective feedback would also account for the retention test means being higher than the posttest means.

All three analyses revealed that the example-only groups required significantly more examples to learn the science than the rule-example groups. This finding also replicated the differences found in the earlier pilot study conducted by the authors. The presentation of the rules reduced the number of examples required to complete the task to almost a minimum. Thus, most Ss in the rule-example groups received only one example for each rule. All of the analyses revealed that the differences on the transfer test were non-significant. This finding was also consistent with the results of the earlier pilot study.

The significant reasoning by treatment interactions found in these studies revealed that the requirement for reasoning may be reduced by providing S with rules in addition to examples. Even though the regression lines cross within the range, the cross-over is spurious since no S received less than ten examples. Therefore, there would be little to gain from assigning high ability Ss to an example-only treatment.

The ability-by-treatment interactions found in these two studies do not replicate those found in the earlier pilot study. In that study, there was no significant reasoning-by-treatment interaction, and the memory-by-treatment interaction was significant. In contrast, the reasoning-by-treatment interactions were significant, and the memory-by-treatment interactions were non-significant in the present studies. In addition to the above differences, the regression line of memory on number of examples for the example-only group had a positive slope in the pilot study, but the corresponding regression lines in the present studies had a negative slope. In other words, as memory scores increased, the number of examples required to complete the task also increased in the pilot study, but in the present study the number of examples decreased as memory increased.

The shifts in the relationships between abilities and treatments from the pilot study to the present studies were apparently due to the revisions made in the task after the pilot study. Therefore, by revising the task, certain variables were manipulated which inadvertently affected the relationships between the abilities and task performance.

On the basis of the results of these studies, it was concluded that a rule-example treatment reduces the requirement for reasoning ability and the number of examples required to reach criterion performance when compared to an example-only treatment. Contrary to the learning-by-discovery hypothesis, an example-only, or discovery, treatment does not produce superior transfer to a higher-order task.

The reasoning-by-treatment interactions obtained in these studies do not conclusively demonstrate that an all-or-none position on learning-by-discovery should be rejected. With this particular task, the rule-example treatment seemed to be superior to the example-only treatment on every performance measure. However, high reasoning $\underline{S}$s in the example-only groups were able to perform as well as the $\underline{S}$s in the rule-example group. The floor effect on the number-of-examples criterion measure made it impossible to determine whether or not an example-only treatment would be more profitable for high-reasoning $\underline{S}$s. By the manipulation of task variables, it is possible to alter the relationship between abilities and task performance. Future aptitude-by-treatment research might be more fruitful if an effort were made to analyze the psychological processes underlying shifts in the relationships between abilities and performance brought about by manipulating task variables, rather than searching for disordinal interactions.

References

Bruner, J. S. The art of discovery. *Harvard Educational Review,* 1961, 31, 21-32.

Bunderson, C. V. Transfer of mental abilities at different states of practice in the solution of concept problems. (Research Bulletin No. RB-67-20) Princeton, N.J.: Educational Testing Service, 1967.

Bunderson, C. V. Aptitude by treatment interaction in designing instruction for a hierarchical learning task. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles, Feb., 1969.

Bunderson, C. V., Olivier, W. P., & Merrill, P. F. The design of an abstract hierarchical learning task for computer-based instructional research. Technical Report No. 2, Contract ONR N00014-67-A-0126-0006, Advanced Research Projects Agency, Computer-Assisted Instruction Laboratory, The University of Texas at Austin, 1971.

Cronbach, L. J. The logic of experiments on discovery. In L. S. Shulman & E. R. Keislar (Eds.), *Learning by discovery: A critical appraisal.* Chicago: Rand McNally, 1966, pp. 77-92.

Cronbach, L. J., & Snow, R. E. Individual differences in learning ability as a function of instructional variables. Final Report, U.S. Office of Education, Contract OEC-4-6-061269-1217. Stanford, Calif.: Stanford University, March 1969.

Dunham, J. L., & Bunderson, C. V. Effect of decision-rule instruction upon the relationship of cognitive abilities to performance in multiple-category concept problems. *Journal of Educational Psychology,* 1969, 60, 2, 121-125.

Dunham, J. L., Guilford, J. P., & Hoepfner, R. Multivariate approaches to discovering the intellectual components of concept learning. *Psychological Review,* 1968, 75, 206-221.

French, J. W., Ekstrom, R. B., & Price, L. A. *Manual kit of reference tests for cognitive factors.* Princeton, N.J.: Educational Testing Service, 1963.

Gagné, R. M., & Paradise, N. E. Abilities and learning sets in knowledge acquisition. *Psychological Monographs,* 1961, 75, No. 518.

14

Merrill, M. D. Transfer effects within a hierarchical task as a function of review and correction on successive parts. Technical Report No. 5, NONR Contract 3985(04). Urbana, Illinois: Training Research Laboratory, 1964.

Shulman, L. S., & Keislar, E. R. (Eds.) *Learning by discovery: A critical appraisal.* Chicago: Rand McNally, 1966.

Wittrock, M. C. The learning by discovery hypothesis. In L. S. Shulman & E. R. Keislar (Eds.), *Learning by discovery: A critical appraisal.* Chicago: Rand McNally, 1966, pp. 33-75.